

グローバルコミュニケーションを担う機械翻訳技術

1. 言語の壁を乗り越える

インターネットの普及が、距離の壁を克服し、海外のあらゆる情報がリアルタイムで入手できるようになり、新たに言葉の壁に注目が集まるようになってきました。

この言語の壁を乗り越え、グローバルコミュニケーションのためのツールとして翻訳ソフトウェアに注目が集まっています。

“グローバルコミュニケーションを担う機械翻訳技術”シリーズでは、3回にわたり翻訳ソフトウェアを支える技術について紹介します。



3種類の機械翻訳技術

翻訳ソフトウェアの命は、翻訳精度が良いことにあります。今回は、この翻訳を行う技術“機械翻訳技術”について説明します。機械翻訳技術は3種類に大別されます。

第一の方式は、“ダイレクト翻訳方式”です(図1(a))。この方式は、Hello. こんにちは。

How are you? お元気ですか? というように、たくさんの例文を覚えておくというものです。しかし、この方法では、記録されている文しか翻訳できません。したがって、翻訳精度の向上のためには、記録している文の数を増加させる必要があります。このため、ダイレクト翻訳方式は、分野を非常に限定した電訳機(例えば、トラベル会話集)などで使用されるだけとなりました。

第二の方式は、“中間言語方式”です(図1(b))。これは、ある文の

意味を解析し、言語によらない共通の概念構造(中間言語)にするものです。つまり、文の“言いたいこと”を抽出してから、それをいろいろな言語で出力するという方式です。この方式は多言語翻訳システムの構築には向いていますが、“元言語 中間言語 目的言語”という2段階の翻訳が必要になり、翻訳精度が良くないという問題があります。

第三の方式は、“意味トランスファ

方式”です。当社の翻訳ソフトウェア“The翻訳™シリーズ”で用いられている翻訳精度の良い方式です。

意味トランスファ方式とは

意味トランスファ方式は、原文意味を解析する際、原文の構造を失わないレベルまでの解析を行い、その結果を対象言語の構造に変換(トランスファ)するというものです。

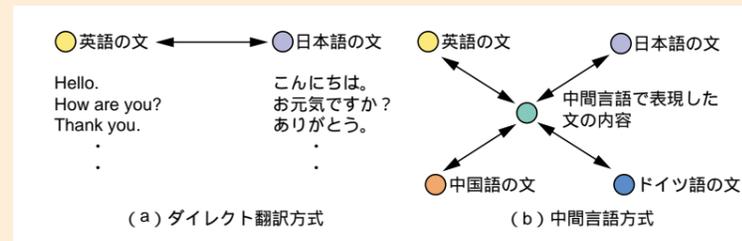


図1. 機械翻訳の方式 機械翻訳技術は3種類に大別され、ダイレクト翻訳方式、中間言語方式にはそれぞれ問題があります。

実際に「He went to New York.」という文を日本語にする例で説明します(図2)。

まず、辞書を引き単語ごとに上記の文を分けます(形態素解析)。次に、この単語と単語の関係(主語 - 述語、修飾 - 被修飾などの関係)を調べ、これをツリー状に構造化します(構文解析)。この段階で原文の構造を保っているため、原文の細かいニュアンスを訳文に反映することができます。更に、英語のことばとことばの関係を日本語のことばの関係に置き換え、単語も英単語から日本語の単語に置き換え(トランスファ)。そして、最後に語順を決めて日本語の文を組み立てます(訳文生成)。

この各過程で、単語の機能や文の構造の分析をいかに的確に行うかが正確な訳文につながります。このための知識が、翻訳辞書に蓄積されています。蓄積されている内容は、語の意味や使われ方です。例えば、「I take a taxi.」、「I take a bath.」、「I take a pen.」という三つの文を例にとり考えると、文の形はすべて同じですが、taxiは乗る物、bathは入る物、penは持てる物という知識が蓄積されているために、それぞれ「タクシーに乗る」、「風呂に入る」、「ペンを取る」と訳し分けることができます(図3)。

機械翻訳技術の課題

しかし、上記のような知識の蓄積があっても、翻訳を100%正確に行うことはできません。機械翻訳が苦手とする例を紹介しましょう。

あいまいな文

例えば、「I saw a girl with a telescope.」という文は、「双眼鏡で少女を見た」と「双眼鏡を持った少女を見た」という二つの解釈があり、

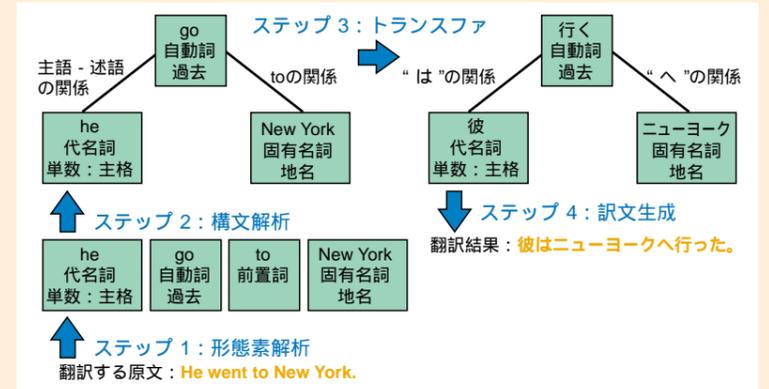


図2. 意味トランスファ方式 原文の細かいニュアンスを伝えることができる優れた方式で、The翻訳™シリーズにも採用されています。

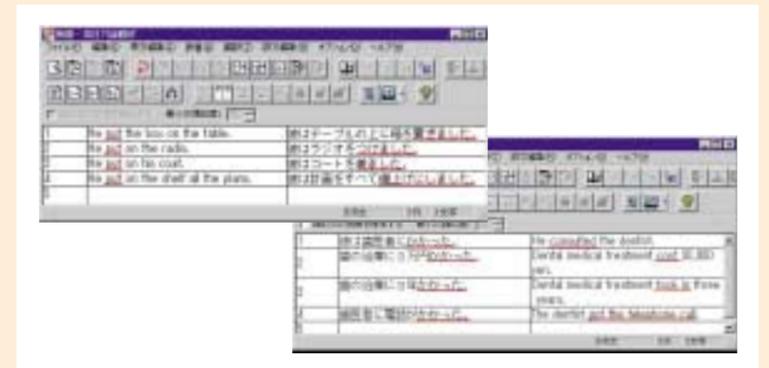


図3. 訳し分けの例 The翻訳™シリーズでは上記のような訳し分けができます。図はThe翻訳™プロフェッショナルの画面です。

この文だけからはどちらが正しいか判定できません。この場合、文脈から状況を判断する必要があります。

省略

日本語の「僕はうなぎだ」の場合、そのまま「I'm an eel.」と訳すのではなく、「僕(が注文するのは)、うなぎ(料理)だ」と語句を補って翻訳する必要があります。どのような語句が省略されているのかを推測するには膨大な知識が必要になります。

上記のような問題は残るものの、

現在の翻訳精度は、文法的に正しい文を文法的に正しい文として8割以上出力することができ、“意味を取る”、“下訳に使う”という使用法であれば十分役立つレベルとなっています。そういったなかで、当社の翻訳ソフトウェア The翻訳™シリーズは、雑誌やインターネットで、翻訳精度が良いとの評価を得ています。

伊藤 悦雄

e-ソリューション社 府中e-ソリューション工場
コンピュータソフトウェア部主務